Collecting and utilising crowdsourced data for numerical weather prediction

Propositions from the meeting held in Copenhagen, 4-5 December 2018

Hintz, Kasper S.; O'Boyle, Katharine; Dance, Sarah L.; Al-Ali, Saja; Ansper, Ivar; Blaauboer, Dick; Clark, Matthew; Cress, Alexander; Dahoui, Mohamed; Darcy, Ronan; Hyrkkanen, Juhana; Isaksen, Lars; Kaas, Eigil; Korsholm, Ulrik S.; Lavanant, Marion; Le Bloa, Gwenaelle; Mallet, Emilie; McNicholas, Conor; Onvlee-Hooimeijer, Jeanette; Sass, Bent; Siirand, Valeria; Vedel, Henrik; Waller, Joanne A.; Yang, Xiaohua

Published in:
Atmospheric Science Letters

DOI:
10.1002/asl.921

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Hintz, K. S., O'Boyle, K., Dance, S. L., Al-Ali, S., Ansper, I., Blaauboer, D., Clark, M., Cress, A., Dahoui, M., Darcy, R., Hyrkkanen, J., Isaksen, L., Kaas, E., Korsholm, U. S., Lavanant, M., Le Bloa, G., Mallet, E., McNicholas, C., Onvlee-Hooimeijer, J., ... Yang, X. (2019). Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4-5 December 2018. *Atmospheric Science Letters*, *20*(7), [UNSP e921]. https://doi.org/10.1002/asl.921

Received: 22 March 2019 | Revised: 16 May 2019 | Accepted: 23 May 2019

DOI: 10.1002/asl.921

RESEARCH ARTICLE

Atmospheric Science Letters RMetS

Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4-5 December 2018

Kasper S. Hintz[1] | **Katharine O'Boyle**[2] | **Sarah L. Dance**[4] | **Saja Al-Ali**[8] | **Ivar Ansper**[9] | **Dick Blaauboer**[5] | **Matthew Clark**[2] | **Alexander Cress**[7] | **Mohamed Dahoui**[6] | **Rónán Darcy**[11] | **Juhana Hyrkkanen**[10] | **Lars Isaksen**[6] | **Eigil Kaas**[12] | **Ulrik S. Korsholm**[1] | **Marion Lavanant**[8] | **Gwenaelle Le Bloa**[8] | **Emilie Mallet**[8] | **Conor McNicholas**[3] | **Jeanette Onvlee-Hooimeijer**[5] | **Bent Sass**[1] | **Valeria Siirand**[9] | **Henrik Vedel**[1] | **Joanne A. Waller**[4] | **Xiaohua Yang**[1]

[1]Danish Meteorological Institute, Copenhagen, Denmark

[2]Met Office, Exeter, UK

[3]Department of Atmospheric Sciences, University of Washington, Seattle, Washington

[4]University of Reading, Reading, UK

[5]Royal Netherlands Meteorological Institute, De Bilt, the Netherlands

[6]European Centre for Medium-Range Weather Forecasts, Reading, UK

[7]Deutsche Wetterdienst, Offenbach, Germany

[8]Météo-France, Paris, France

[9]Estonian Environment Agency, Tallinn, Estonia

[10]Finnish Meteorological Institute, Helsinki, Finland

[11]Met Èireann, Dublin, Ireland

[12]University of Copenhagen, Niels Bohr Institute, Copenhagen, Denmark

Correspondence
Kasper S. Hintz, Danish Meteorological Institute, Copenhagen, Denmark.
Email: kah@dmi.dk

Funding information
Innovation Fund Denmark, Grant/Award Number: 5189-00042B; UK EPSRC, Grant/Award Number: EP/P002331/1

Abstract

In December 2018, the Danish Meteorological Institute organised an international meeting on the subject of crowdsourced data in numerical weather prediction (NWP) and weather forecasting. The meeting, spanning 2 days, gathered experts on crowdsourced data from both meteorological institutes and universities from Europe and the United States. Scientific presentations highlighted a vast array of possibilities and progress being made globally. Subjects include data from vehicles, smartphones, and private weather stations. Two groups were created to discuss open questions regarding the collection and use of crowdsourced data from different observing platforms. Common challenges were identified and potential solutions were discussed. While most of the work presented was preliminary, the results shared suggested that crowdsourced observations have the potential to enhance NWP. A common platform for sharing expertise, data, and results would help crowdsourced data realise this potential.

KEYWORDS

citizen science, crowdsourcing, data collection, opportunistic data, quality control, third-party data

1 | INTRODUCTION

Within the atmospheric sciences, "crowdsourced" data is a relatively new term. While the term crowdsourcing was initially defined by Howe (2006) as outsourcing an act to the general public, this definition is no longer restricted to traditional tasks being outsourced. Today, crowdsourcing is more than outsourcing data collection to the general public. Instead, crowdsourcing embraces new data sources, data storage, quality control and utilisation, which requires standard methods and a common terminology.

Direct and indirect observations from non-conventional sources are being investigated for use in the atmospheric sciences. Examples of data sources include Personal Weather Stations (PWSs) (Bell *et al.,* 2013, 2015; Clark *et al.,* 2018), smartphones (Kim *et al.,* 2015; McNicholas and Mass, 2018; Price *et al.,* 2018; Hintz *et al.,* 2019), vehicles (Anderson *et al.,* 2012; Mahoney and O'Sullivan, 2013) and communication networks (Zinevich *et al.,* 2009).

Muller *et al*. (2015) provided a comprehensive review of "crowdsourcing" efforts in the atmospheric sciences. Since this review was published, new advancements have been made with crowdsourced datasets. Some of the most recent advancements include the collection and quality-control of atmospheric pressure observations from smartphones (Kim *et al.,* 2015, 2016; Madaus and Mass, 2017; McNicholas and Mass, 2018; Price *et al.,* 2018; Hintz *et al.,* 2019). Kim *et al*. (2016) was the first to apply machine-learning methods to bias correct smartphone pressure observations (SPOs). McNicholas and Mass (2018) demonstrated an efficient machine-learning approach to SPO bias correction that benefited from non-meteorological smartphone sensor data. Clark *et al*. (2018) examined the use of PWSs and made considerable progress in the quality control and use of such data. Examples of successful assimilation of such observations into operational Numerical Weather Prediction (NWP) models are currently few and far between. The integration of observations from PWSs into the NOAA Meteorological Assimilation Data Ingest System (NCEP, 2019) dataset is an early example. Also, in the U.S., the utility of PWSs has been of increasing interest for forecasts of severe convection (Madaus *et al.,* 2014; Carlaw *et al.,* 2015; Sobash and Stensrud, 2015; Gasperoni *et al.,* 2018).

A meeting on the use of crowdsourced data in NWP and weather forecasting was held in Copenhagen 4-5th December 2018 at the Danish Meteorological Institute (DMI), with two main purposes. First, to gather experts within the topics of crowdsourcing and create a network of people working on the subject, and second to discuss common issues encountered with crowdsourced data and how these can be addressed. Researchers from both universities and meteorological institutes attended the meeting, whose experience spanned a variety of subjects, including SPOs, PWSs, vehicular data, and citizen weather reports. The first day was allocated for presentations from the participants, followed by plenary discussion. The second day was allocated for discussions, starting with a sketch of ongoing activities at Institutions and Universities. Two working groups were created who reviewed current research topics for various data sources and data formats. The purpose of this article is to document the propositions and recommendations from the meeting and to inform peers of ongoing activities.

2 | SCIENTIFIC PRESENTATIONS

C. McNicholas (University of Washington) discussed how measurements of atmospheric pressure could be efficiently retrieved from smartphones and subsequently bias-corrected. Results from a testbed Android app, uWx, revealed that inaccuracies in smartphone location and sensor internal filtering contributed to poor data quality. Correcting these issues facilitated the retrieval of pressure change without the need for post-processing/quality control. Using a machine learning approach, smartphone pressures were bias-corrected to account for large uncertainties in smartphone elevation (McNicholas and Mass, 2018). For each smartphone, a random forest was trained on auxiliary sensor/GPS data to predict and correct pressure errors. On average, bias correction reduced pressure errors by ~ 80%. During post-processing, fewer than 20% of smartphone pressure were discarded. In a real-world case-study bias-corrected smartphone pressures improved analyses and 1-hour forecasts of altimeter setting, 2-minute temperature, and 2-minute dewpoint.

K. S. Hintz (DMI) first presented a study on wind measurements from smartphones, in which the surface roughness length was estimated from the measured horizontal turbulence. In another work, more than 6 million SPOs were collected over 7 weeks through a software development kit installed in a third-party mobile app. These observations were quality controlled and assimilated with 3D-Var in the DMI HARMONIE NWP system (Hintz *et al.,* 2019; Yang et al., 2017). A decrease of bias and no change in root mean squared error was found for a simulation period of nearly 2 months. Examples showing that raw observations can depict current weather was given.

X. Yang (DMI) presented the construction idea behind the operational COntinuous Mesoscale Ensemble Prediction System (COMEPS) (Yang et al., 2017a, 2017b) at DMI used for a routine weather forecasts, which generates a 2.5 km grid resolution, 25 member, Rapid Update Cycle (RUC) like EPS forecast with an hourly update using time lagging. Currently, a proto-type ensemble nowcasting system applying the COMEPS approach is in development, targeting subhourly cycling of the high-resolution nowcasting system assimilating high frequency observation data such as radar and crowdsourced data. One of the novel system components in COMEPS is the time-lagged 3D-Var analysis on overlapped observation windows, which appears especially beneficial in nowcasting applications with variational data assimilation, as the setup appears to have better potential to address observation error correlation in time and space, as well as the issue of model spin-up in connection with frequent assimilation cycling.

A. Cress (Deutscher Wetterdienst, DWD), presented the activities of DWD concerning crowdsourced data applications and their use in the local DWD data assimilation system. Within the Fleet Weather Maps Project FloWKar, a collaboration between DWD and the German car manufacturer AUDI AG has been established, to investigate to what extent future environmental observations from vehicle sensors can be combined with existing data sources to improve nowcasting and warnings and therefore make a contribution to the security of future autonomous driving. A complete real-time weather conceptual framework has been established, focusing on the flow and processing of high resolution measurements and weather products and the development of corresponding forecasts. A fast data exchange is followed by quality control according to weather service standards and smart aggregation strategies, integrating all available data into a real-time weather map. Aiming for fast weather forecasting, a data assimilation cycle with a 5-minute update rate is necessary; therefore, an ultra-rapid data assimilation method is proposed. A real-world application employs the high resolution project observations in a 5-minute assimilation cycle for the regional operational weather model COSMO-D2, focusing on the model performance optimisation near the surface and its predictions along road sections in Germany, where the current observation network is not dense enough. First results, comparing car measurements, nearby weather stations and model analysis and forecasts were presented.

E. Mallet and S. Al Ali (Météo France) first gave a brief overlook of crowdsourcing activities at Météo-France. Those activities focus on the use of human observations from “expert” non-professional observers and “citizen” observers, and automated observations collected from PWSs, agricultural networks and connected vehicles. Then, the presentation focused on two ongoing projects: (1) The first action concerns the crowdsourcing module in Météo-France's mobile application that allows mobile users to report the observed weather and to post pictures of the sky. The module provides, without access restriction, a simple entry of almost twenty phenomena to the users, who in their turn will select the observed phenomena and report the observed weather condition on a regular basis. Based on this module, more than 10,000 observations are collected daily, and more than 40,000 in high-risk situations. Visualisation of crowdsourced data is already available to forecasters, and the next step is to feed it to operational databases in order to expand its possible uses. (2) The second action concerns the potential use of vehicle observations for meteorological applications which is the subject of a partnership between Météo-France and Continental. The aim is to infer weather (precipitation and low visibility) and road conditions (dry, wet, slick) at a particular location in time, through the analysis of vehicle data elements (temperature, wiper and headlight statuses, velocity, and the activation of ABS and ESP systems). The experimental campaign started in November 2016 and is still ongoing. The fleet consists of hundreds of vehicles, transmitting data through a connected dongle. Data filtering and quality checking routines were developed, and vehicle observations were evaluated against meteorological data. Machine learning classification algorithms were developed, using data from meteorological observation merging products as references for hydro-meteor discrimination and visibility. The preliminary results were promising and also showed the need to combine multiple parameters in order to successfully derive weather observations.

K. O'Boyle (Met Office) presented how The Met Office view crowdsourcing as distinct from citizen science (see section 4). There is a long history of citizen science at the Met Office. The Weather Observations Website (WOW) (wow.metoffice.gov.uk) is the Met Office citizen science portal. WOW has global reach, and is a platform for anybody to submit, share and display their weather observations, either manually or by connecting a PWS using APIs. WOW data is being trialled in nowcasting applications, but is not yet assimilated into NWP. Investigations into other opportunistic observations are ongoing at the Met Office, including collecting data from vehicles.

M. Clark (Met Office) presented on an automated quality control and gridding process for citizen science data. There has been a focus on Met Office WOW data from PWSs to create high resolution surface analyses. Parameter values from each WOW site are constrained to have the same long-term mean as neighbouring official sites, but are otherwise allowed to vary freely, as it is assumed that shorter-term, temporary deviations are the signature of genuine small scale features which are worth retaining in the analysis. A series of case studies have shown that there is value in this approach.

S.L. Dance (University of Reading) gave an overview of the DARE: Data Assimilation for the REsilient City project. This is a UK Engineering and Physical Sciences Research Council (EPSRC) Senior Fellowship in Digital Technology for Living with Environmental Change. The vision for the project is to use “datasets of opportunity”, such as CCTV and vehicle observations, alongside scientific observing networks, such as satellite data (Mason *et al.,* 2018) to improve predictions of urban natural hazards such as flooding and high impact weather. There are many potential benefits of such data, including the availability of large numbers of inexpensive observations, in areas where there are people but there may be few sources of scientific observation data. For example, (1) air traffic management reports have potential to provide observations of temperature inversions in the boundary layer (Mirza *et al.,* 2016, 2019). (2) In many locations around the world, the population has access to smartphones, but ground-based scientific observations are sparse. Furthermore, there are a number of issues in collecting 'datasets of opportunity' for use in assimilation. These include the need for metadata such as time and location in order to carry out the assimilation, versus data protection for the data provider, who may be a private individual. Other issues include data ownership, intermittency, heterogeneity, data provenance and large data volumes. In order to use such observations in data assimilation, there needs to be an understanding of natural variability in urban areas (where many of these data originate) and the variability that can be resolved by a prediction model (e.g., Waller *et al.,* 2014; Janjić *et al.,* 2017). This was discussed further in the next talk by J.A. Waller.

J. A. Waller (University of Reading) presented on the potential to measure temperatures in urban areas using vehicles. Issues related to the assimilation of crowdsourced data were discussed; in particular, the need to understand the data inhomogeneity and natural variability of observation urban areas in order to understand the observation uncertainties. Collaborative work with the UK Met Office, is assessing the potential of temperature observations recorded by vehicles. The preliminary findings showed that the data collection method was not reliable for collecting large temperature data sets. Furthermore, for the initial data sets collected, it was shown that temperature measurements had a negative correlation with the speed of the vehicle. It was concluded that a new data collection technique was required, and a more detailed study was vital before the benefits of assimilating vehicle temperatures could be assessed.

D. Blaauboer (KNMI and EUMETNET) presented shortly the KNMI-activities in the domain of crowdsourcing. These include participation in the WOW project of UK Met Office, application of car data (temperature sensor, wiper data), smartphone data, damage reporting app (to report weather impacts by the public), wind data from hot air balloons. EUMETNET, the grouping of 31 European National MetServices, recognised the emerging availability and application opportunities of crowdsourced data and the Internet of Things among many of its members. Therefore EUMETNET has organised a few dedicated workshops on this subject with the aim to bring experts in this field together, to foster networking and possibly create a platform or programme in near future to develop common applications to the benefit of all.

M. Dahoui (ECMWF) presented an overview of the importance on in-situ data in global NWP. It was shown that there are data gaps in the surface observations received at ECMWF and the potential and challenges for using crowdsourced data to fill these gaps were described. Also, it was stressed that crowdsourced data can be important for verification purposes. A denser network is useful to detect small scale features and rapid changes of the atmosphere, so observations have also the potential to improve the forecast verification aspects leading to a better understanding of model performance. The usage of crowdsourced observations is however very challenging. It was suggested that data collection and pre-processing needs a collaborative effort between NWP centres through coordination of the WMO, the industry and the private sector to improve and unify standards and to agree on best practices. A common and shared use of operationally managed data hubs (such as the MetOffice Weather Observation Website) is a cost-effective solution to manage the diversity of data sources and formats. A good understanding of the error characteristics of the observations is necessary to allow proper data selection and error specification. This requires a comprehensive and standardised description of metadata. Quality control, bias correction and blacklist management require unique identification of a reporting station which makes anonymous reports of less interest to NWP data assimilation unless technological solutions are available to anonymously identify the data or perform most of the quality control and bias correction near the data origin. Legal aspects related to privacy and data usage are also essential to clarify before the operational use of such observations.

3 | OVERVIEW OF ACTIVITIES

During the meeting, it became clear that there are many activities on-going, with opportunities for collaboration. Table 1 list activities, status and considerations for participating institutions together with ZAMG and Met Norway who agreed to share their current activities. It is seen that especially work with data from private weather stations is an active field of research at many institutions.

4 | CHALLENGES AND SOLUTIONS

The presentations and discussions identified several common challenges, and some solutions were proposed during the discussion sessions. These follow below:

TABLE 1 Overview of ongoing and considered activities at each participating institute and institutes that was not present but approved to be included

Institution	Current activities	Activity status	Considerations	Contact persons
KNMI	WOW-NL	Research, operational	Pollution measurements	Marijn De Haij
Met Office	WOW-UK, social media, cars, voluntary observations	Research, operational	User reports, 5G network	Katharine O'Boyle
DMI	SPO, PWS	Research	User reports, webcam	Kasper Hintz
FMI	User reports	Operational	SPO	Juhana Hyrkkanen
Mètèo France	Cars, PWS, user reports	Research		Èmilie mallet
University of Reading	Cars, CCTV, WOW	Research	Buses	Sarah Dance
ECMWF	Monitoring Progress	Research		Mohamed Dahoui
DWD	Cars, PWS	Research	SPO, user reports	Alexander cress
Met Èireann	Voluntary observations	Resarch, operational	PWS	Ronan Darcy
ESTEA	App in development			Ivar Ansper
University of Washington	SPO	Research		Conor McNicholas
ZAMG	Trusted spotter network, Austrian weather observer	Operational		Thomas Krennert
Met Norway	PWS (Netatmo)	Research, operational		Roger Randriamampianina

The activities at University of Reading are only including activities within the DARE (data assimilation for the REsilient city) project.

I. Terminology is not agreed upon in the community. A common vocabulary needs to be established to facilitate future collaboration. The term “crowdsourced data” is used differently within the community, and there is no agreement what this term covers and what not. Often crowdsourced data is used as a collective term, for example, citizen-science and third-party data, which is how the term will be treated in this report, though with a recognition that a more precise definition is desirable.
  i. The Met Office suggested a terminology that clearly separates citizen-science data and crowdsourced data, and also attempts to define associated terms:
    a. Citizen-science data: Information obtained from a group of people who are invited to participate in a data collection process.
    b. Crowdsourced data: Information derived from a group of people without their explicit involvement in the data collection process.
    c. Opportunistic data: Information derived from non-meteorological sensors or weather sensitivities.
    d. Third-party data: Data collected by a third-party organisation using meteorological sensors.

  However, some similarities are expected between third-party data and the other groups. For example PWS observations might be classified as both third-party data and crowdsourced data.
  ii. ECMWF proposed four main categories of ‘crowdsourced’ data; private and third party, automated amateur weather stations, smart connected devices (mobile phones and vehicles), and human reporting of the current weather, relating each of these to the ease of utility in NWP.

In the terminology proposed by the Met Office (i), there is a clear separation between citizen-science data and ‘crowdsourced’ data, wherein the ECMWF proposal (ii) the term ‘crowdsourced’ data is a collective term. It is recommended that authors define their usage of these terms.

II. Obtaining useful crowdsourced data may involve collaboration with commercial entities, such as manufacturers of PWSs or vehicles. In some cases, collaborations of this kind mark a step change in the way universities and meteorological institutes have previously operated. For professional use, crowdsourced data needs to be as unprocessed as possible when received. Working in collaboration with manufacturers may enable this. Some of the workshop participants have built successful collaborations with commercial entities, taking a "virtuous circle" approach, whereby data is provided by a manufacturer, and in return the meteorological institution provides forecast data or quality controlled observational data. It is crucial that intellectual property rights and data ownership are clear and agreed upon before starting collaborations.

III. Law based restrictions on storage of personal data lead to a need to de-personalise crowdsourced data, which can lead to "black boxes". Metadata can be used to help characterise the error of crowdsourced observations, and for bias correction, but the legal constraints regarding privacy and personal data can limit the collection of such metadata. Hence, metadata vs privacy is one issue that must be considered when collecting observations. DMI have invested in legal expertise and are open to sharing the information obtained with the community. This is mainly related to the European GDPR regulation (European Union, 2018).

IV. New data sources can potentially produce more observations than current NWP models can realistically handle. New methods, such as those suggested by Dr. X. Yang (DMI), will need to be considered. Tendencies of parameters are not commonly assimilated into NWP; a change in approach may be required to extract maximum value from crowdsourced observations.

   i. It was discussed that data streaming could be a way of handle the amount of observations in future, such that, in operational systems, observations that come in are utilised and then thrown away. This may seem somewhat provocative to some as the NWP community are often used to store data for an extended time. However, it was agreed that data streaming could perhaps be only realistic solution currently to overcome issues with data volume. Also near-real communication could perhaps be easier to implement with a streaming approach.

   ii. The scale of crowdsourced observations, any reference network, and NWP models will all be different. To make them comparable, methods to deal with multiscale comparisons are required for example, filtering or superobbing.

Further, other themes seemed to be well established. There was a general agreement that crowdsourced data can provide useful observations in areas otherwise devoid of observations. It was discussed whether stationary platforms (e.g., PWS) are easier to implement in existing systems than moving platforms (e.g., vehicles, SPOs). In general, stationary platforms are believed to be easier to bias-correct than moving platforms. Also, new data sources should be seen to supplement conventional observation networks rather than a replacement, as trusted observations are required as a reference for new data sources. A nested platform of reference may be a good way of organising networks in the future, for example, SYNOPs used as a reference for the quality control of PWS data, and PWS then used as a more dense reference dataset for observations from mobile platforms.

5 | CONCLUSIONS AND RECOMMENDATIONS

Much of the work presented at the workshop was at an early, exploratory stage, and many questions remain unanswered. However, a general set of conclusions were drawn from the discussion. Crowdsourced observations are potentially useful for NWP, and are undoubtedly useful for verification and forecasting. Use of crowdsourced observations in nowcasting, or post-processing, is perceived to be easier and less demanding than in NWP data assimilation. There is still much work to do before crowdsourced observations can widely be ingested into NWP models.

It was agreed upon that there is a sliding scale between 'crowdsourced' or "passive" data collection, where an individual's involvement is limited, and "citizen science" or "active" data collection where the individual is explicitly involved. It is generally thought that the lesser degree of interaction required by the participant the higher the volume of data that can be collected. It is not clear if either of the two are of superior quality.

Further, the following recommendations are made. An organised community of those involved in crowdsourcing activities would be beneficial. EUMETNET would provide a good forum for this, however, such a forum should not be restricted to European countries. This forum could be a simple, independent, platform accessible via a website. Regarding vocabulary, it would be beneficial for the community to agree on common terminology related to crowdsourcing. To realise the full potential of crowdsourced data for NWP, issues of data quality, privacy, and availability will need to be addressed. Data quality could be enhanced by prioritising the collection of accurate metadata. Privacy issues should be addressed to determine if, how, and when unique identifiers can be retrieved for quality control purposes. Lastly, efforts to expand crowdsourced datasets by disseminating data operationally and working with private industry should be encouraged.

ACKNOWLEDGEMENTS

The authors wish to thank DMI meteorologist Ole Kristensen for giving an operational aspect on the use of crowdsourced data. The contribution of S.L Dance and J.A. Waller was funded in part by UK EPSRC Grant No. EP/P002331/1 Data Assimilation for the Resilient City (DARE).

ORCID

Kasper S. Hintz https://orcid.org/0000-0002-6835-8733

REFERENCES

Anderson, A.R.S., Chapman, M., Drobot, S.D., Tadesse, A., Lambi, B., Wiener, G. and Pisano, P. (2012) Quality of mobile air temperature and atmospheric pressure observations from the 2010 development test environment experiment. *Journal of Applied Meteorology and Climatology*, 51(4), 691–701.

Bell, S., Cornford, D. and Bastin, L. (2013) The state of automated amateur weather observations. *Weather*, 68(2), 36–41.

Bell, S., Cornford, D. and Bastin, L. (2015) How good are citizen weather stations? Addressing a biased opinion. *Weather*, 70(3), 75–84.

Carlaw, L.B., Brotzge, J.A. and Carr, F.H. (2015) Investigating the impacts of assimilating surface observations on high-resolution forecasts of the may 15, 2013 tornado event. *Electronic Journal of Severe Storms*, 10(2), 1–34.

Clark, M.R., Webb, J.D.C. and Kirk, P.J. (2018) Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations. *Meteorological Applications*, 25(3), 472–492.

European Union (2018). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. (accessed June 16, 2018).

Gasperoni, N.A., Wang, X., Brewster, K.A. and Carr, F.H. (2018) Assessing impacts of the high-frequency assimilation of surface observations for the forecast of convection initiation on April 3, 2014 within the dallasfort worth test bed. *Monthly Weather Review*, 146(11), 3845–3872.

Hintz, K.S., Vedel, H. and Kaas, E. (2019) Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation. *Meteorological Applications*, 2019, 1–14. https://doi.org/10.1002/met.1805.

Howe, J. (2006). Crowdsourcing: A definition. http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html. (accessed September 22, 2018).

Janjić, T., Bormann, N., Bocquet, M., Carton, J.A., Cohn, S.E., Dance, S.L., Losa, S.N., Nichols, N.K., Potthast, R., Waller, J.A. and Weston, P. (2017) On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144 (713), 1257–1278.

Kim, N.-Y., Kim, Y.-H., Yoon, Y., Im, H.-H., Choi, R.K.Y. and Lee, Y.H. (2015) Correcting air-pressure data collected by mems sensors in smartphones. *Journal of Sensors*, 2015, 1–10.

Kim, Y.-H., Ha, J.-H., Yoon, Y., Kim, N.-Y., Im, H.-H., Sim, S. and Choi, K.Y.R. (2016) Improved correction of atmospheric pressure data obtained by smartphones through machine learning. *Computational Intelligence and Neuroscience*, 2016, 12. https://doi.org/10.1155/2016/9467878.

Madaus, L.E., Hakim, G.J. and Mass, C.F. (2014) Utility of dense pressure observations for improving mesoscale analyses and forecasts. *Monthly Weather Review*, 142(7), 2398–2413.

Madaus, L.E. and Mass, C.F. (2017) Evaluating smartphone pressure observations for mesoscale analyses and forecasts. *Weather and Forecasting*, 32(2), 511–531.

Mahoney, W.P. and O'Sullivan, J.M. (2013) Realizing the potential of vehicle-based observations. *Bulletin of the American Meteorological Society*, 94(7), 1007–1018.

Mason, D.C., Dance, S.L., Vetra-Carvalho, S. and Cloke, H.L. (2018) Robust algorithm for detecting floodwater in urban areas using synthetic aperture radar images. *Journal of Applied Remote Sensing*, 12, 12–20.

McNicholas, C. and Mass, C. (2018) Smartphone pressure collection and bias correction using machine learning. *Journal of Atmospheric and Oceanic Technology*, 35, 523–540.

Mirza, A.K., Ballard, S.P., Dance, S.L., Maisey, P., Rooney, G.G. and Stone, E.K. (2016) Comparison of aircraft-derived observations with in situ research aircraft measurements. *Quarterly Journal of the Royal Meteorological Society*, 142(701), 2949–2967.

Mirza, A.K., Ballard, S.P., Dance, S.L., Rooney, G.G. and Stone, E.K. (2019) Towards operational use of aircraft-derived observations: a case study at London heathrow airport. *Meteorological Applications*, 2019, 1–14. https://doi.org/10.1002/met.1782.

Muller, C., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A. and Leigh, R. (2015) Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology*, 35(11), 3185–3203.

NCEP (2019). Madis. https://madis.ncep.noaa.gov. (accessed January 17, 2019).

Price, C., Maor, R. and Shachaf, H. (2018) Using smartphones for monitoring atmospheric tides. *Journal of Atmospheric and Solar-Terrestrial Physics*, 174, 1–4.

Sobash, R.A. and Stensrud, D.J. (2015) Assimilating surface mesonet observations with the enkf to improve ensemble forecasts of convection initiation on may 29, 2012. *Monthly Weather Review*, 143 (9), 3700–3725.

Waller, J.A., Dance, S.L., Lawless, A.S., Nichols, N.K. and Eyre, J.R. (2014) Representativity error for temperature and humidity using the met Office high-resolution model. *Quarterly Journal of the Royal Meteorological Society*, 140(681), 1189–1197.

Yang, X., Feddersen, H., Sass, B. and Sattler, K. (2017a) *Construction of a Continuous Mesoscale Eps with Time Lagging and Assimilation on Overlapping Windows*, ALADIN-HIRLAM Newsletter, Vol 8, pp. 112–118.

Yang, X., Stig, B., Dahldom, M., Sass, B., Zhuang, S., Amstrup, B., Petersen, C., Nielsen, K., Nielsen, N.W. and Mahura, A. (2017b) *NEA, the Operational Implementation of HARMONIE 40h1.1 at DMI*, ALADIN-HIRLAM Newsletter, Vol 8, pp. 104–111.

Zinevich, A., Messer, H. and Alpert, P. (2009) Frontal rainfall observation by a commercial microwave communication network. *Journal of Applied Meteorology and Climatology*, 48(7), 1317–1334.

How to cite this article: Hintz KS, O'Boyle K, Dance SL, *et al.* Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4-5 December 2018. *Atmos Sci Lett*. 2019;20:e921. https://doi.org/10.1002/asl.921